

Emergent Collaboration in Social Purpose Games*

Robert P. Gilles[†]

Lina Mallozzi[‡]

Roberta Messalli[§]

September 2021

Abstract

We study a class of non-cooperative aggregative games—denoted as *social purpose games*—in which the payoffs depend separately on a player’s own strategy (individual benefits) and on a function of the strategy profile which is common to all players (social benefits) weighted by an individual benefit parameter. This structure allows for an asymmetric assessment of the social benefit across players.

We show that these games have a potential and we investigate its properties. We investigate the payoff structure and the uniqueness of Nash equilibria and social optima. Furthermore, following the literature on partial cooperation, we investigate the leadership of a single coalition of cooperators while the rest of players act as non-cooperative followers. In particular, we show that social purpose games admit the emergence of a stable coalition of cooperators for the subclass of *strict* social purpose games. Due to the nature of the partial cooperative leadership equilibrium, stable coalitions of cooperators reflect a limited form of farsightedness in their formation.

As a particular application, we study the *tragedy of the commons game*. We show that there emerges a single stable coalition of cooperators to curb the over-exploitation of the resource.

Keywords: Partial cooperation; Leadership equilibrium; Potential games; Aggregative games; Tragedy of the Commons.

JEL classification: C71, D71

*We thank Stef Tijs, Emiliya Lazarova and Subhadip Chakrabarti for their inspirational leadership in the partially cooperative development of the equilibrium concepts used in this paper.

[†]Management School, The Queen’s University of Belfast, Riddel Hall, 185 Stranmillis Road, Belfast, BT9 5EE, UK. Email: r.gilles@qub.ac.uk

[‡]Department of Mathematics and Applications, University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy. E-mail: mallozzi@unina.it

[§]Department of Economic and Statistical Science, University of Naples Federico II, Complesso Monte Sant’Angelo 21, 80125 Naples, Italy. E-mail: robertamessalli@gmail.com

1 Introduction: Endogenous partial cooperation in games

In this paper, following the literature of additively separable aggregative games (Corchón, 1994; Dubey et al., 2006) and in line with the asymmetry considered in McGuinty and Milam (2013), we introduce a class of non-cooperative games—denoted as *Social Purpose Games*—for which the payoff of each player depends separately on his own strategy and on a function of the strategy profile, the aggregation function, which is the same for all players, weighted by an individual benefit parameter which enlightens the asymmetry, between agents, of public benefit. The two parts of the payoff function represent respectively the individual and the social benefits.

The purpose of this paper is to investigate the properties of and relationship between Nash equilibria and social optima in social purpose games. Furthermore, we show that for a subclass of these social purpose games there emerges a stable coalition of cooperators founded on the presence of a limited form of farsightedness in the corresponding definition of stability.

The main feature of social purpose games is that they capture the tension between social benefits and individual payoffs. As such this class of social purpose games includes implementations of the tragedy of the commons, pollution abatement games, and public good provision games. In most of these social purpose games, individual optimality leads players to underutilise social benefits in favour of direct individualistic payoffs. We show that, indeed, this is a general feature of social purpose games, exemplified by the social suboptimality of the Nash equilibria. If we interpret the strategies in these social purpose games to represent a chosen level of “effort”, the total Nash equilibrium generated level of effort is socially suboptimal. We argue that partial collaboration through the emergence of a stable coalition of cooperators, partially abates the suboptimality of the resulting effort levels. Therefore, this fundamental insight holds for a substantial class of aggregative games.

For our analysis we introduce two subclasses of social purpose games. In a *regular* social purpose game, all payoff functions are continuous and (weakly) concave. Regularity of a social purpose game guarantees the existence of the main equilibrium concepts. A social purpose game is *strict* if the payoff structures are strictly convex, guaranteeing uniqueness of the main equilibrium concepts.

Jensen (2010) already pointed out that classes of aggregative games can be characterised through certain forms of potential functions. We show here that social purpose games admit a weighted potential function. This is a substantially stronger property than Jensen showed for certain classes of aggregative games. Furthermore, the existence of a weighted potential leads to the question regarding the relationship between the maximisers of the potential function and the Nash equilibria of these games. We show that for a special class of social purpose games the set of Nash equilibria actually coincides with the set of potential maximisers.

Concerning the Nash equilibria of social purpose games, we show a wide range of properties. Under standard continuity properties, any social purpose game admits at least one Nash equilibrium, including the class of regular social purpose games. For the class of *strict* social purpose games the Nash equilibrium is unique due to the strict convexity of the corresponding payoff structures.

Similarly, under these same conditions, social purpose games generally admit at least one social optimum, while for the subclass of strict social purpose games the social optimum is unique. The

fundamental tension between social and individual benefit leads to the natural conclusion that, as stated above, for a large subclass of social purpose games, the Nash equilibrium levels of effort are socially suboptimal. The next goal of our paper is to investigate how partial cooperation can alleviate tension between socially optimal and equilibrium levels of effort.

Partial cooperation in social purpose games Social purpose games have a specific structure that allows the emergence of stable partial cooperation among players. Our investigation is rooted in the work on partial cooperation in a range of types of non-cooperative games by Chander and Tulkens (1997), Mallozzi and Tijs (2008, 2009, 2012) and Chakrabarti et al. (2011, 2018). In this literature, one considers the formation of a single “coalition of cooperators” that collectively determines a joint strategy to maximise its collective payoffs—being the sum of the individual members’ payoffs. This is akin to cartel formation in oligopolistic market games (D’Aspremont et al., 1983) and international treaty writing in the context of environmental abatement situations (Diamantoudi and Sartzetakis, 2006; Kwon, 2006; Olmstead, 2014). There are two natural behavioural hypotheses that can be considered in this context.

First, one can assume that the coalition of cooperators acts as a single player under standard best response rationality. The resulting stable strategy profiles are referred to as *partial cooperative equilibria*, which existence can be established under relatively mild conditions (Chakrabarti et al., 2011, 2018). This avenue is not investigated in our analysis, since in many applications the coalition of cooperators would not have a standard position in the corresponding decision making processes; in most natural applications, the coalition of cooperators is found to have a position of *leadership* in a decision-making hierarchy.

The hypothesis, that the coalition of cooperators assumes a leadership position as a first mover in relation to the non-cooperating players in the game, was seminally proposed by von Stackelberg (1934) and Chander and Tulkens (1997). This idea has been developed further as the notion of a *partial cooperative leadership equilibrium* (PCLE) in the cited literature on partial cooperation in general non-cooperative games. It is clear that the leadership position of the coalition of cooperators gives it an advantage in comparison to the standard partial cooperative equilibrium payoff. Nevertheless, this leadership position seems natural and has been observed in the context of social purpose games, in particular for public good provision games, pollution abatement games, and (oligopolistic) market games.

In our study of partial cooperation in social purpose games, we limit ourselves to the investigation of PCLE under the formation of a coalition of cooperators. In particular, we can show that under relatively mild conditions there exists a unique PCLE in a social purpose game. This conception, therefore, allows us the introduction of farsightedness in coalition formation in these games. Namely, we assume that if a coalition of cooperators emerges in a social purpose game, it assumes a leadership position and there naturally emerges a unique PCLE.

This affords us with the possibility to consider the emergence of a “stable” coalition of cooperators. The applied notion of stability in coalition formation is founded on the von Neumann-Morgenstern standards (von Neumann and Morgenstern, 1947), which considers two partial stability properties: (1) *Internal stability*—every cooperator will obtain a lower payoff upon leaving the coalition of

cooperators, and (2) *External stability*—every non-cooperator receives a lower payoff upon joining the existing coalition of cooperators.

We show for a subclass of the strict social purpose games there indeed emerge stable coalitions of cooperators that abate the social suboptimality of the corresponding equilibrium effort levels.

Relationship to the literature An aggregative game is founded on the hypothesis that each payoff function depends on the corresponding player's strategy as well as on some aggregation of all selected strategies. Classical examples of aggregation are the unweighted sum and the mean. The concept of aggregative games goes back to Selten (1970), who considers as aggregation function the summation of all the players' strategies. Later, this concept has been studied in the case of other aggregation functions and it has been generalised to the concept of quasi-aggregative games. For this we refer to Vives (1990), Corchón (1994), Cornes and Hartley (2005), Dubey et al. (2006), De Miguel and Xu (2009), Jensen (2010), Acemoglu and Jensen (2013), and Mallozzi and Messalli (2017). In this literature, there are many games that present an aggregative structure: among them, we mention Cournot and Bertrand games, patent races, models of contests of fighting and model with aggregate demand externalities.

In additively separable aggregative games, each payoff function is a sum of a function that depends on an aggregation of strategies and a function that depends on player's own strategy. The model of additively separable aggregative games appeared in literature, among others, in the context of International Environmental Agreements (IEA), studying the formation of stable IEA in the case in which each country's choice variable is emission and then extending the results to the dual case, i.e., the case where the choice variable is abatement effort (Diamantoudi and Sartzetakis, 2006).

Also public good provision games are in the context of additively separable aggregative games (Bergstrom et al., 1986) where each player consumes a certain amount of a private good and donates a certain other amount to the supply of the public good. Thus, the payoff function of each player turns to depend not only on the quantity of private good that he consumes but also on all contributions to the public good made by all individuals. McGuinty and Milam (2013) investigate the impact of asymmetry in a voluntary public goods environment by proposing an improved design that explicitly isolates individual incentives, without assuming a dominant strategy.

Structure of the paper In Section 2 of this paper we investigate thoroughly the quintessential social purpose game, namely a standard implementation of the tragedy of the commons (Hardin, 1968). We show that the effects of the over-exploitation of the commons can be mitigated by partial cooperation among the players and that such cooperation can be stable. Stable partial cooperation is shown to mitigate the over-exploitation of the commons.

In Section 3 we introduce the class of general social purpose games and identify some relevant sub-classes. We show that social purpose games in general admit a weighted potential and that for a sub-class of these games the potential maximisers coincide with the set of Nash equilibria. We discuss social optima and their properties, showing in particular that for the class of regular social purpose games there is the expected relationship between Nash equilibrium and socially optimal strategies, referring to the over-exploitation of the commons.

Section 4 introduces the notion of partial cooperation in social purpose games. We define the notion of a partial cooperative leadership equilibrium and prove its existence for regular social purpose games. Subsequently, we introduce a stability concept in the formation of partial cooperation and show that it is natural to expect that strict social purpose games admit the emergence of stable partial cooperation. This is fully developed for an application to quadratic payoff functions.

2 A motivating case: The tragedy of the commons

The “tragedy of the commons” refers to the classical problem of the overuse and exploitation of a common resource through free-riding in a non-cooperative setting. Traditionally the common resource referred to a common tract of land in a medieval village for the grazing of cattle owned by the village’s peasants. If peasants freely access and use the land, a situation of over-exploitation arises, resulting in the depletion of the commons for collective use (Lloyd, 1833).¹

The tragedy of the commons was formulated in game-theoretic terms by Hardin (1968) and has been considered a totemic reference in many contributions to the social and biological sciences (Frischmann et al., 2019). Hardin (1968) only considered the non-cooperative case of unlimited and free extraction from the commons. Ostrom (1990) challenged Hardin’s reductionism by pointing out that in many social situations the commons was and still is successfully governed through the application of institutional and behavioural solutions. Gilles et al. (2020) introduce a mathematical model of an institutional solution on managing the commons from a public good perspective based on Ostrom’s insights.

Here we pursue a third perspective on the management of the commons by investigating the endogenous emergence of a coalition of cooperators that collectively regulate their extraction from the commons, while non-members of this coalition of cooperators selfishly extract. We can show that the extraction from the commons under such *partial cooperation* significantly improves collective wealth generation and welfare. We illustrate the benefits from endogenous partial cooperation in the classical tragedy of the commons game by considering a simple example.

A non-cooperative extraction game Suppose that there is a finite commonly owned resource that has a total size of one (1). There are $n \in \mathbb{N}$ with $n \geq 3$ users of this resource, who are individually free to extract any benefit from the common resource with the understanding that any future benefit from the resource would be limited by the extent of today’s usage. Hence, future use is based on the remainder of the common resource at the conclusion of today’s collective extraction.

This results in a standard non-cooperative game $\langle N, (S_i)_{i \in N}, (\pi_i)_{i \in N} \rangle$ where $N = \{1, 2, \dots, n\}$ is the set of users and $S_i = [0, 1]$ is the set of individual extraction levels. We apply a standard payoff function that considers benefits from current usage and the utility of future usage with equal weight

¹The tragedy of the commons has significant appeal and application in our contemporary global economy, referring to contemporary issues such as the exploitation of natural resources—including fish stocks, fresh water sources, mining of ores, and oligopolistic commodity markets—as well as the global environmental conditions. The tragedy of the commons is known as one of the most fundamental examples of a *social dilemma*.

for every $i \in N$. Hence,

$$\pi_i(x) = x_i \left(1 - \sum_{j \in N} x_j \right) = x_i (1 - X_N) \quad \text{for every } x \in S = \prod_{j \in N} S_j = [0, 1]^n \subset \mathbb{R}^n.$$

We denote by $X_T = \sum_{i \in T} x_i$ the total extraction of a coalition $T \subseteq N$ with the convention that $X_\emptyset = 0$. Note that $(1 - X_N)$ therefore represents the total size of the commons left for future usage.²

Nash equilibria and social optima We summarise the resulting Nash equilibrium³ and the unique social (Pareto) optimum⁴ outcomes in the following table:

	x_i	X_N	π_i	$\sum \pi_i$
Nash equilibrium	$\frac{1}{n+1}$	$\frac{n}{n+1}$	$\frac{1}{(n+1)^2}$	$\frac{n}{(n+1)^2}$
Social optimum	$\frac{1}{2n}$	$\frac{1}{2}$	$\frac{1}{4n}$	$\frac{1}{4}$

We observe here the well-known conclusion that there significant over-extraction in the Nash equilibrium, which reduces welfare uniformly for all users. Indeed, we note that $X_N^{\text{NE}} > X_N^{\text{SO}}$ as well as $\sum \pi_i^{\text{NE}} < \sum \pi_i^{\text{SO}}$ for all $n \geq 3$.

Partial cooperation in the tragedy of the commons Next we consider a hybrid of best response rationality and socially optimal decision-making by allowing users to collaborate to extract in a collectively rational fashion. In particular, consider that a coalition of cooperators $C \subset N$ decides collectively about their extraction rate, while the non-collaborating users $j \in N \setminus C$ act according to best response rationality. Furthermore, we assume that the coalition of cooperators C assumes a Stackelberg leadership position and acts as a first-mover. Hence, users in C determine their coordinated extraction rates taking into account the best responses of all non-cooperators $j \in N \setminus C$. This is referred to as a *partial cooperative leadership equilibrium* (PCLE) in Chakrabarti et al. (2011, 2018).

Suppose that the coalition of cooperators has size $|C| = m < n$. In a PCLE, non-cooperators $j \in N \setminus C$ optimise their payoffs given X_C as well as $(x_k)_{k \in N \setminus C, k \neq j}$. This results for the non-cooperator $j \in N \setminus C$ in solving

$$\max_{0 \leq x_j \leq 1} \pi_j(x_j, x^{NC}, x^C) = x_j (1 - X_C - X_{NC})$$

where $X_C = \sum_{i \in C} x_i$ and $X_{NC} = \sum_{h \in N \setminus C} x_h$. Solving this simultaneously for all $j \in N \setminus C$, given X_C , we conclude that the best response for $j \in N \setminus C$ is given by

$$x_j(X_C) = \frac{1 - X_C}{n - m + 1} \quad \text{and} \quad \sum_{j \in N \setminus C} x_j(X_C) = \frac{(n - m)(1 - X_C)}{n - m + 1} \quad (1)$$

²We remark that $1 - X_N < 0$ refers to the destruction of the commons, resulting in negative payoffs for all users.

³A Nash equilibrium is a strategy profile $x^{\text{NE}} \in S$ in which all users do a best response to what other users extract from the commons, i.e., $\pi_i(x^{\text{NE}}) = \max_{x_i \in S_i} \pi_i(x_i, x_{-i}^{\text{NE}})$.

⁴A social optimum is a strategy profile $x^{\text{SO}} \in S$ that maximizes the collective welfare, i.e., $\sum_{i \in N} \pi_i(x^{\text{SO}}) = \max_{x \in S} \sum_{i \in N} \pi_i(x)$.

Next, the coalition of cooperators C collectively determine their collective extraction, given the optimal decisions of all non-cooperators, by solving

$$\max_{x^C \in [0,1]^C} \sum_{i \in C} \pi_i \left(x^C, (x_j(X_C))_{j \in N \setminus C} \right) = X_C \left(1 - X_C - \sum_{j \in N \setminus C} x_j(X_C) \right)$$

This results in the conclusion that the coalition of cooperators C solves

$$\max_{x^C \in [0,1]^C} X_C \cdot \frac{1 - X_C}{n - m + 1} \quad (2)$$

leading to the conclusion that the optimal collective extraction is $X_C^L = \frac{1}{2}$. The resulting PCLE can be summarised as

$$x_i^L = x_C^L(m) = \frac{1}{2m} \quad \text{for } i \in C \quad (3)$$

$$x_j^L = x_{NC}^L(m) = \frac{1}{2(n - m + 1)} \quad \text{for } j \in N \setminus C \quad (4)$$

resulting in $X_C^L(m) = \frac{1}{2}$, $X_{NC}^L(m) = \frac{n-m}{2(n-m+1)}$ and $X_N^L(m) = 1 - \frac{1}{2(n-m+1)}$. Clearly, the resulting PCLE outcomes form a hybrid between Nash equilibrium and social optimum extraction. The resulting PCLE payoffs reflect this as

$$\pi_i^L = \pi_C^L(m) = \frac{1}{4m(n - m + 1)} \quad \text{for } i \in C \quad (5)$$

$$\pi_j^L = \pi_{NC}^L(m) = \frac{1}{4(n - m + 1)^2} \quad \text{for } j \in N \setminus C \quad (6)$$

Stable partial cooperation The tragedy of the commons as formulated here allows for the consideration of stability in the process of partial cooperation under leadership. We call a coalition of cooperators C *stable* if no non-cooperator $j \in N \setminus C$ would like to join the coalition C and no member $i \in C$ would like to leave the coalition of cooperators. Note that this introduces a model of coalition formation that is founded on one-step farsightedness on part of all users in this tragedy of the commons situation.

Formally, we say that a coalition of cooperators $C \subseteq N$ is stable if

$$\pi_{NC}^L(m) \geq \pi_C^L(m+1) \quad \text{as well as} \quad \pi_C^L(m) \geq \pi_{NC}^L(m-1) \quad (7)$$

After some calculus we conclude that a coalition of cooperators C is stable if and only if

$$\frac{m+1}{n-m+1} \geq \frac{n-m+1}{n-m} \quad \text{as well as} \quad \frac{n-m+2}{n-m+1} \geq \frac{m}{n-m+2}$$

It can be verified that there are many (n, m) pairs that satisfy these inequalities. In particular, this holds for $n = 8$ and $m = 5$. We verify this for these particular sizes of the coalition of cooperators for population size $n = 8$. The next table summarises the main resulting PCLE strategies and extraction rates for all sizes $2 \leq m < n = 8$ of the coalition of cooperators.

m	x_C^L	x_{NC}^L	X_N	π_C^L	π_{NC}^L
2	$\frac{1}{4}$	$\frac{1}{14}$	$\frac{13}{14}$	$\frac{1}{56}$	$\frac{1}{196}$
3	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{11}{12}$	$\frac{1}{72}$	$\frac{1}{144}$
4	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{9}{10}$	$\frac{1}{80}$	$\frac{1}{100}$
5*	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{7}{8}$	$\frac{1}{80}$	$\frac{1}{64}$
6	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{5}{6}$	$\frac{1}{72}$	$\frac{1}{36}$
7	$\frac{1}{14}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{56}$	$\frac{1}{16}$

From the table of the computed PCLE, we conclude that indeed $m^* = 5$ is a stable size of the coalition. Indeed, $\pi_C^L(5) = \frac{1}{80} > \pi_{NC}^L(4) = \frac{1}{100}$ as well as $\pi_{NC}^L(5) = \frac{1}{64} > \pi_C^L(6) = \frac{1}{72}$.

The analysis of the PCLE in the setting of the tragedy of the commons shows that there emerges self-regulation of the extraction from the commons through partial cooperation under a limited farsighted behavioural rationality. In this paper we explore a more general class of non-cooperative games that exhibits this feature. This is developed over the following sections.

3 Social purpose games

In the theory of non-cooperative games, a *normal form game* is an interactive decision situation that is represented as a list $G = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ with

- $N = \{1, \dots, n\}$ is a given finite set of players, where $n \in \mathbb{N}$ is the number of players;
- for every $i \in N$, S_i is a non-empty strategy or action set for player i , where we define $S = \prod_{i \in N} S_i$ as the set of all strategy tuples in G , and;
- for every $i \in N$, $\pi_i: S \rightarrow \mathbb{R}$ is the payoff function of player i .

A *Nash equilibrium* in the game $G = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ is a strategy tuple $x^* \in S$ such that for every $i \in N$: $\pi_i(x^*) \geq \pi_i(x_{-i}^*, x_i)$ for every alternative strategy $x_i \in S_i$, where the tuple of strategies of all players except i is denoted by $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \prod_{j \neq i} S_j$.

We now introduce a social purpose game as a non-cooperative normal form game with a specific payoff structure.

Definition 3.1 A *social purpose game* is a list $\Gamma = \langle N, \bar{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ that defines an associated non-cooperative game $G_\Gamma = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ with

- $N = \{1, \dots, n\}$ is a finite player set, where the number of players is $n \in \mathbb{N}$ such that $n \geq 2$;
- $\bar{Q} > 0$ is a positive number such that every $i \in N$: $S_i = [0, \bar{Q}]$, endowed with the standard Euclidean topology;
- $H: \mathbb{R} \rightarrow \mathbb{R}$ is a function, and;
- For every $i \in N$: $\alpha_i > 0$ is a parametric weight and $h_i, g_i: [0, \bar{Q}] \rightarrow \mathbb{R}$ are functions such that for every $x = (x_1, \dots, x_n) \in S = [0, \bar{Q}]^N$ player i 's payoffs are given by

$$\pi_i(x) = \alpha_i H \left(\sum_{j=1}^n h_j(x_j) \right) - g_i(x_i) \quad (8)$$

Throughout we refer to Γ as a “game” in the same sense as a non-cooperative game defined above and use it synonymously with G_Γ .

All payoffs in a social purpose game depend separately on player i 's own strategy, through the term $g_i(x_i)$ that represents an individual cost of player i of executing strategy x_i , and on a common term $H\left(\sum_{i=1}^n h_i(x_i)\right)$ that represents a common social benefit for all players. Therefore, the function H is naturally interpreted as a *social benefit function*, while for every player $i \in N$ the function g_i represents an *individual cost function*.

The functions $(h_i)_{i \in N}$ can naturally be interpreted as assessments of individual contributions in the aggregated contribution, represented as $\sum_{i \in N} h_i$. If h_i is the identity function for all $i \in N$, we arrive at the standard utilitarian formulation of aggregation. On the other hand, our formulation allows for a wide range of different representations of aggregation through selections of the functions H and $(h_i)_{i \in N}$. We refer to the functions H and $(h_i, g_i)_{i \in N}$ as the *constituting functions* of the social purpose game Γ .

Furthermore, the payoff of player $i \in N$ depends on the common benefit, weighted by the individual benefit parameter, $\alpha_i > 0$, that measures the importance that player i gives to the common benefit H in comparison with the individual cost g_i . This allows for asymmetry in the assessment of the common benefit across players as individual contributors to the common benefit. For ease of the following analysis, we assume that all players in N are ranked according to their preference for the generated social benefit, i.e., such that $0 < \alpha_1 \leq \dots \leq \alpha_n$.

A social purpose game Γ is clearly an aggregative separable game—applying the insights from [Acemoglu and Jensen \(2013\)](#). The class of social purpose games includes the classes of various well-known aggregative games such as *pollution abatement games* and, more generally, *public good provision games*, which are both characterised by $\alpha_i = 1$ for all players $i \in N$.

Remark 3.2 We note that the tragedy of the commons situation considered in Section 2 is indeed a social purpose game. This is made explicit by noting that the payoff function of the tragedy of the commons can be written in the required form through a monotonic transformation. Indeed, taking the (natural) logarithm of her payoff function we can represent the payoff function of user $i \in N$ as $\log \pi_i(x) = \log(1 - \sum_{i \in N} x_i) - (-\log x_i)$. Hence, this corresponds to a social purpose payoff function determined by $\alpha_i = 1$, $h_i(x_i) = x_i$, $g_i(x_i) = -\log x_i$ and $H(x) = \log(1 - x)$. \blacklozenge

Classes of social purpose games In the discussion of social purpose games we have not imposed any properties on the constituting functions H and $(h_i, g_i)_{i \in N}$. It is natural to require these functions to satisfy certain properties. In particular, we consider the case in which h_i are the identity functions, which refers to a class of social purpose games that include the public good provision games.

This simplification introduces several interesting classes of social purpose games. For these classes of games, we are able to compare the aggregate Nash equilibrium strategies and the aggregate social optimum strategies, as well as the size of the individual contributions in the Nash equilibrium and the social optimum.

Definition 3.3 Let $\Gamma = \langle N, \bar{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ be a social purpose game, generating the normal form representation G_Γ .

- (i) The social purpose game Γ is called **regular** if
- all functions h_i , $i \in N$, are the identity functions $h_i(x_i) = x_i$ and all functions H and $(g_i)_{i \in N}$ are continuously differentiable;
 - the common benefit function H is continuously differentiable, increasing and concave—implying that its derivative H' is weakly decreasing on X , and;
 - the individual cost function g_i is continuously differentiable, increasing and convex in x_i for every $i \in N$ —implying that its derivative g'_i is weakly increasing for every $i \in N$.
- (ii) The social purpose game Γ is called **strict** if Γ is regular and, additionally, for every $i \in N$ the individual cost function g_i is twice differentiable and strictly convex—implying that $g''_i > 0$.

Regularity of social purpose games refers to the concavity of the payoff functions. This implies that regular social purpose games describe a situation in which players make a contribution to a common goal or resource and that these contributions are subject to weakly decreasing returns to scale, individually as well as socially. This class of social purpose games captures a very large number of relevant applications, including environmental problems, the provision of public goods and extractive situations from a common resource.

Strict social purpose games impose that the payoff structure is strictly concave. This additional property implies that one can think of these games as extractive situations from a common resource that is subject to strictly decreasing marginal returns. One can also refer to these strict social purpose games as “extraction games” to emphasise the most common and obvious application of this class of games. In particular, we note that the tragedy of the commons is an extraction game that can be represented as a strict social purpose game.

3.1 Potentials and Nash equilibria

In the following proposition we show that each social purpose game is a weighted potential game with weights α_i for all players $i \in N$ (Monderer and Shapley, 1996).

Proposition 3.4 *Every social purpose game $\Gamma = \langle N, \bar{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ is a weighted potential game for weight vector $\alpha = (\alpha_1, \dots, \alpha_n)$. The corresponding α -potential is given by*

$$P(x) = H\left(\sum_{i=1}^n h_i(x_i)\right) - \sum_{i=1}^n \frac{1}{\alpha_i} g_i(x_i) \quad (9)$$

Proof. We prove the result checking that the definition of weighted potential game holds true. For every $x_i, y_i \in X_i$ and $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in [0, \bar{Q}]^{n-1}$ we have that:

$$\begin{aligned}
\pi_i(y_i, x_{-i}) - \pi_i(x_i, x_{-i}) &= \\
\alpha_i H(h_i(y_i) + \sum_{j \neq i} h_j(x_j)) - g_i(y_i) - \alpha_i H(h_i(x_i) + \sum_{j \neq i} h_j(x_j)) + g_i(x_i) &= \\
\alpha_i \left(H(h_i(y_i) + \sum_{j \neq i} h_j(x_j)) - \frac{1}{\alpha_i} g_i(y_i) - H(h_i(x_i) + \sum_{j \neq i} h_j(x_j)) + \frac{1}{\alpha_i} g_i(x_i) \right) &= \\
\alpha_i \left(H(h_i(y_i) + \sum_{j \neq i} h_j(x_j)) - \frac{1}{\alpha_i} g_i(y_i) - \sum_{j \neq i} \frac{1}{\alpha_j} g_j(x_j) - H(h_i(x_i) + \sum_{j \neq i} h_j(x_j)) + \right. & \\
\left. + \frac{1}{\alpha_i} g_i(x_i) + \sum_{j \neq i} \frac{1}{\alpha_j} g_j(x_j) \right) = \alpha_i (P(y_i, x_{-i}) - P(x_i, x_{-i})). &
\end{aligned}$$

This concludes the proof of the assertion. ■

Properties of Nash equilibria of social purpose games We say that social purpose game Γ admits Nash equilibria if the associated normal form game G_Γ has Nash equilibria. Using established insights from potential game theory, we can prove quite straightforwardly the existence of Nash equilibrium for the class of social purpose games.

Proposition 3.5 Consider a social purpose game Γ .

- (a) If all constituting functions H and $(h_i, g_i)_{i \in N}$ are continuous, then Γ admits at least one Nash equilibrium $x^{\text{NE}} = (x_1^{\text{NE}}, \dots, x_n^{\text{NE}}) \in [0, \bar{Q}]^N$.
- (b) If the social purpose game Γ is strict such that H is non-constant, then Γ admits a unique Nash equilibrium.
- (c) If all constituting functions H and $(h_i, g_i)_{i \in N}$ are continuously differentiable, then every interior Nash equilibrium $x^{\text{NE}} = (x_1^{\text{NE}}, \dots, x_n^{\text{NE}}) \in (0, \bar{Q})^N$ is a solution to the set of equations given by

$$H' \left(\sum_{j=1}^n h_j(x_j) \right) h'_i(x_i) = \frac{g'_i(x_i)}{\alpha_i} \quad \text{for every } i \in N. \tag{10}$$

Proof. The hypotheses that H is continuous and h_i and g_i are continuous for every $i \in N$, imply that the potential function H is continuous on $[0, \bar{Q}]^N$. By compactness of $[0, \bar{Q}]^N$, there exists a maximum of P and, thus, since for a weighted potential game all potential maximisers are Nash equilibria (Monderer and Shapley, 1996), i.e., $\arg \max P \subseteq \text{NE}_\Gamma$, where NE_Γ is the set of Nash equilibria of G_Γ , there exists at least one Nash equilibrium in G_Γ .

To show assertion (b), for any $s = \sum_{i=1}^n x_i$ let $x_i(s) \geq 0$ be the solution of $\alpha_i H'(s) = g'_i(x_i(s))$. Since g'_i is strictly monotone, we find $x_i(s) = (g'_i)^{-1}(\alpha_i H'(s))$. Since Γ is a strict social purpose game and $(g'_i)^{-1}$ is increasing, the map $s \rightarrow x_i(s)$ is continuous and decreasing on $[0, n\bar{Q}]$.

We define the map $F: s \in [0, n\bar{Q}] \mapsto \sum_{i=1}^n x_i(s) \in [0, n\bar{Q}]$, which is continuous and decreasing on $[0, n\bar{Q}]$, since $\sum_{i=1}^n x_i(0) > 0$. Furthermore, under our assumptions $H'(0) > 0$, implying that F has a unique fixed point $s^* \in [0, n\bar{Q}]$, i.e., $s^* = \sum_{i=1}^n x_i(s^*)$, corresponding to the unique Nash equilibrium.

To show assertion (c), consider an interior Nash equilibrium $x^{\text{NE}} = (x_1^{\text{NE}}, \dots, x_n^{\text{NE}}) \in (0, \bar{Q})^N$. Then clearly x^{NE} satisfies the first order conditions of the Nash best response optimisation problem given by

$$x_i^{\text{NE}} \in \arg \max_{0 < x_i < \bar{Q}} \pi_i(x_i, x_{-i}^{\text{NE}}).$$

Noting that

$$\frac{\partial \pi_i}{\partial x_i} \pi_i(x) = \alpha_i H' \left(\sum_{j=1}^n h_j(x_j) \right) h'_i(x_i) - g'_i(x_i),$$

we arrive immediately at the conclusion that assertion (c) is valid. \blacksquare

Remark 3.6 In the case in which the constituting functions are continuously differentiable, any interior Nash equilibrium is a solution of the system of first order conditions given by $\nabla P(x_1, \dots, x_n) = 0$. This implies that every interior Nash equilibrium is a stationary point of the potential P . This allows the relatively easy computation of the Nash equilibria of a social purpose game.

Note here that given equations in Proposition 3.5(c) lead to the conclusion that for all $i, j \in N$:

$$\frac{g'_i(x_i^{\text{NE}})}{\alpha_i h'_i(x_i^{\text{NE}})} = \frac{g'_j(x_j^{\text{NE}})}{\alpha_j h'_j(x_j^{\text{NE}})}$$

Hence, each of these fractions represents a *shadow price* of the individual marginal benefit in terms of the common marginal benefit.

If h_i is the identity function for all $i \in N$, this shadow price is simplified to $\rho_i = \frac{g'_i(x_i^{\text{NE}})}{\alpha_i}$. These shadow prices can be used to interpret the tensions between individual costs and benefits in social purpose game situations. \blacklozenge

Clearly, by the definition of weighted potential game we have that $P^{\text{max}} \subseteq \text{NE}(\Gamma)$, where $P^{\text{max}} = \arg \max_{[0, \bar{Q}]^N} P(x)$ is the set of α -potential maximizers and $\text{NE}(\Gamma)$ the set of the Nash equilibria of the game G_Γ . For a special class of social purpose games, it is possible to characterise the set of the potential maximizers as the set of the Nash equilibria.

Proposition 3.7 *Suppose that the social benefit function is linear, i.e., $H(t) = at + b$ with $a, b \in \mathbb{R}$. Then it holds that $P^{\text{max}} = \text{NE}(\Gamma)$.*

Proof. Under the linear assumption, the potential function is:

$$P(x) = a \left(\sum_{i=1}^n h_i(x_i) \right) + b - \sum_{i=1}^n \frac{g_i(x_i)}{\alpha_i}$$

We must prove the inclusion $\text{NE}(\Gamma) \subseteq P^{\max}$. Suppose that we have a Nash equilibrium of the game Γ denoted by (x_1^*, \dots, x_n^*) such that for any $i \in N$ and any $x_i \in [0, \bar{Q}]$:

$$\alpha_i H \left(\sum_{i=1}^n h_i(x_i^*) \right) - g_i(x_i^*) \geq \alpha_i H \left(h_i(x_i) + \sum_{j \neq i} h_j(x_j^*) \right) - g_i(x_i)$$

that is equivalent to

$$H \left(\sum_{i=1}^n h_i(x_i^*) \right) - \frac{g_i(x_i^*)}{\alpha_i} \geq H \left(h_i(x_i) + \sum_{j \neq i} h_j(x_j^*) \right) - \frac{g_i(x_i)}{\alpha_i}$$

This is under the linear assumption equivalent to

$$a \left(\sum_{i=1}^n h_i(x_i^*) \right) + b - \frac{g_i(x_i^*)}{\alpha_i} \geq a \left(h_i(x_i) + \sum_{j \neq i} h_j(x_j^*) \right) + b - \frac{g_i(x_i)}{\alpha_i}$$

Summing up all the inequalities, we conclude that for all $i \in N$:

$$na \left(\sum_{i=1}^n h_i(x_i^*) \right) + nb - \sum_{i=1}^n \frac{g_i(x_i^*)}{\alpha_i} \geq a \left(\sum_{i=1}^n h_i(x_i) \right) + (n-1)a \left(\sum_{i=1}^n h_i(x_i^*) \right) + nb - \sum_{i=1}^n \frac{g_i(x_i)}{\alpha_i}$$

leading to

$$a \left(\sum_{i=1}^n h_i(x_i^*) \right) - \sum_{i=1}^n \frac{g_i(x_i^*)}{\alpha_i} \geq a \left(\sum_{i=1}^n h_i(x_i) \right) - \sum_{i=1}^n \frac{g_i(x_i)}{\alpha_i}$$

Taking into account that this holds for all $x \in [0, \bar{Q}]^N$, we conclude that $P(x^*) \geq P(x)$ for all x , implying that $x^* \in P^{\max}$. This shows the assertion. \blacksquare

Let us remark that the linearity of H does not fully characterise the equality proved in Proposition 3.7. Indeed, with reference to Example 3.18, there exist social purpose games in which H is not linear and the potential maximizers coincide with the Nash equilibria of the game.

Furthermore, we conjecture that the assertion of Proposition 3.7 holds for the larger class of social purpose games in which H is concave and piecewise linear. The following example shows this for a specific case.

Example 3.8 Consider a social purpose game Γ with two players $n = 2$ such that $\bar{Q} = 1$ and a payoff structure determined by $\alpha_1 = \alpha_2 = 1$, $h_1(x) = h_2(x) = x$, the common payoff function given by

$$H(x) = \begin{cases} x & \text{for } x \in [0, 1] \\ 1 & \text{for } x > 1 \end{cases}$$

and $g_1(x) = g_2(x) = x^2$, resulting in $\pi_1(x_1, x_2) = H(x_1 + x_2) - x_1^2$ and $\pi_2(x_1, x_2) = H(x_1 + x_2) - x_2^2$.

We note that the unique Nash equilibrium of Γ is given by $(\frac{1}{2}, \frac{1}{2})$, which coincides with the unique potential maximiser. \blacklozenge

Remark 3.9 We remark in relation to Proposition 3.7 that for every strict social purpose game Γ that admits a *unique* Nash equilibrium, it holds that there is a unique potential maximiser which is the unique Nash equilibrium. \blacklozenge

3.2 Social optima

We say that the strategy profile $x^{\text{SO}} \in [0, \bar{Q}]^N$ is a *social optimum* of the social purpose game $\Gamma = \langle N, \bar{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ if

$$x^{\text{SO}} \in \arg \max_{x \in [0, \bar{Q}]^N} \sum_{i \in N} \pi_i(x) \quad (11)$$

The next proposition addresses the existence of social optima in social purpose games. For the following analysis it is useful to introduce the *social welfare function* $W: [0, \bar{Q}]^N \rightarrow \mathbb{R}$ where

$$W(x) = \frac{1}{A} \sum_{i=1}^n \pi_i(x) = H \left(\sum_{i=1}^n h_i(x_i) \right) - \frac{\sum_{i=1}^n g_i(x_i)}{A} \quad (12)$$

with $A = \sum_{i \in N} \alpha_i$. Clearly, social optima maximise the function W over $[0, \bar{Q}]^N$. This is used in the proof of the next proposition.

Proposition 3.10 Consider a social purpose game Γ .

- (a) If the common benefit function H and all individual functions $(h_i, g_i)_{i \in N}$ are continuous, then there exists a social optimum $x^{\text{SO}} = (x_1^{\text{SO}}, \dots, x_n^{\text{SO}}) \in [0, \bar{Q}]^N$.
- (b) Suppose that Γ satisfies the following conditions:
 - The common benefit function H is increasing and concave;
 - For every $i \in N$: h_i is concave, and;
 - For every $i \in N$: g_i is strictly convex.

Then there exists a unique social optimum $x^{\text{SO}} \in [0, \bar{Q}]^N$ for Γ .

Proof. Under the stated assumptions, it is obvious that the social welfare function W is continuous on $[0, \bar{Q}]^N$. Hence, applying the Weierstrass theorem, W has a maximum on $[0, \bar{Q}]^N$, which corresponds to a social optimum, showing the assertion.

To show assertion (b), note that from the imposed conditions on H and $(h_i, g_i)_{i \in N}$, the constructed social welfare function W is concave on $[0, \bar{Q}]^N$. Moreover, since g_i is strictly convex for every player $i \in N$, it follows that the social welfare function W is strictly concave. Hence, W has a unique maximiser in $[0, \bar{Q}]^N$, corresponding to the unique social optimum for Γ . \blacksquare

Corollary 3.11 The following properties hold:

- (a) Every regular social purpose game admits a Nash equilibrium as well as a social optimum.
- (b) Every strict social purpose game admits a unique social optimum.

Remark 3.12 If all constituting functions are continuously differentiable, the social optimum solves $\nabla W(x_1, \dots, x_n) = 0$ implying that for every $i \in N$:

$$H' \left(\sum_{j=1}^n h_j(x_j) \right) h'_i(x_i) = \frac{g'_i(x_i)}{A}$$

This compares to the shadow prices derived in Remark 3.6, indicating the well-established underinvestment for the common benefit in situations like represented by social purpose games. \blacklozenge

Remark 3.13 For every strategy tuple $x = (x_1, \dots, x_n) \in X$: $P(x) \leq W(x)$. \blacklozenge

Remark 3.14 The generic component of the social optimum x_i^{SO} does not depend on α_i for every $i \in N$. \blacklozenge

Comparing Nash equilibria and social optima in social purpose games For the class of social purpose games, the comparison of the Nash equilibria and social optima is usually very informative. The next example considers a very simple payoff structure in which the difference between these two concepts is maximal on the strategy set $[0, \bar{Q}]$.

Example 3.15 Consider a regular social purpose game $\Gamma = \langle N, \bar{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ with the following payoff structure:

- $H \left(\sum_{i \in N} h_i(x_i) \right) = \sum_{i \in N} x_i$ is a standard utilitarian common benefit function, where $h_i(x_i) = x_i$ is selected as the identity function for every $i \in N$
- $g_i(x_i) = x_i$ for every $i \in N$, and;
- $\alpha_i < 1$ for every $i \in N$ such that $A = \sum_{i \in N} \alpha_i > 1$.

In this example, for any $i \in N$ we can write $\pi_i(x) = \alpha_i \sum_{j \neq i} x_j + (\alpha_i - 1)x_i$ and we see that the unique Nash equilibrium is given by $x_i^{\text{NE}} = 0$ for all $i \in N$.

On the other hand, the social optimum is identified by maximising the social welfare function

$$\sum_{i \in N} \pi_i(x) = A \sum_{i \in N} x_i - \sum_{i \in N} x_i = (A - 1) \sum_{i \in N} x_i$$

which has a unique maximum identified as being given by $x_i^{\text{SO}} = \bar{Q}$ for every $i \in N$. \blacklozenge

In regular social purpose games, the collective contributions in Nash equilibrium are always lower than the collective contributions required for a social optimum. This refers to the commonly accepted property that is exhibited in public good provision situations.

Proposition 3.16 Suppose that Γ is a regular social purpose game. Then for every Nash equilibrium x^{NE} and every social optimum x^{SO} of G_Γ it holds that

$$\sum_{i=1}^n x_i^{\text{NE}} \leq \sum_{i=1}^n x_i^{\text{SO}}. \quad (13)$$

Proof. Let us suppose by contradiction that $\sum_{i=1}^n x_i^{\text{NE}} > \sum_{i=1}^n x_i^{\text{SO}}$. Since H' is decreasing in $\sum_{i=1}^n x_i$, then $H'(\sum_{i=1}^n x_i^{\text{NE}}) \leq H'(\sum_{i=1}^n x_i^{\text{SO}})$. Since Remarks 3.6 and 3.12 hold true, it follows that for any given $i \in N$: $\frac{g'_i(x_i^{\text{NE}})}{\alpha_i} \leq \frac{g'_i(x_i^{\text{SO}})}{A}$, where $A = \sum_{j=1}^n \alpha_j > 0$.

Since, for any $i \in N$, $0 < \alpha_i < A$ and g'_i is positive, it follows that

$$\frac{g'_i(x_i^{\text{NE}})}{A} < \frac{g'_i(x_i^{\text{NE}})}{\alpha_i} \leq \frac{g'_i(x_i^{\text{SO}})}{A}. \quad (14)$$

Since g'_i is increasing in x_i , we obtain that $x_i^{\text{NE}} \leq x_i^{\text{SO}}$. Hence, we have shown that $x_i^{\text{NE}} \leq x_i^{\text{SO}}$ for any $i \in N$. This, in turn, implies that $\sum_{i=1}^n x_i^{\text{NE}} \leq \sum_{i=1}^n x_i^{\text{SO}}$, contradicting the hypothesis that $\sum_{i=1}^n x_i^{\text{NE}} > \sum_{i=1}^n x_i^{\text{SO}}$. This contradiction proves the assertion. \blacksquare

Remark 3.17 We can easily generalise the previous result to the case in which functions h_i for all $i \in N$ are not the identity functions, instead assuming that $\frac{g'_i(\cdot)}{h'_i(\cdot)}$ are increasing functions in x_i for all players $i \in N$.

This assumption simply means that an increase of player i 's strategy has a greater impact on the individual cost $g_i(x_i)$ than on the common benefit $H(\sum_i h_i(x_i))$ through which player i 's contribution is assessed. \blacklozenge

We also note that, if we assume that the common benefit function H is convex, we lose the property stated in Proposition 3.16.

Example 3.18 We consider the non-regular social purpose game Γ defined by $n = 2$, $\bar{Q} = 1$ and payoff functions π_1 and π_2 given by $\alpha_1 = \alpha_2 = \frac{1}{2}$, $H(t) = 2t^2 - 4t$ and $g_1 = g_2 = 0$; then

$$\pi_1(x_1, x_2) = \pi_2(x_1, x_2) = (x_1 + x_2)^2 - 2(x_1 + x_2). \quad (15)$$

We identify two Nash equilibria $\{(0, 0), (1, 1)\}$ that are also social optima. So, the statement in Proposition 3.16 is no longer true. \blacklozenge

We next compare each component of a Nash equilibrium with the corresponding component of a social optimum. First, we consider an example of a strict social purpose game in which there are players $i \in N$ with $x_i^{\text{NE}} < x_i^{\text{SO}}$ as well as players $j \in N$ with $x_j^{\text{NE}} > x_j^{\text{SO}}$.

Example 3.19 Let $\delta > 0$. Consider a strict social purpose game Γ_δ with $N = \{1, 2\}$ and $\bar{Q} = 1$ with

$$\pi_1(x_1, x_2) = \alpha_1 \log(x_1 + x_2 + \delta) - x_1^{1+\delta} \quad (16)$$

$$\pi_2(x_1, x_2) = \alpha_2 \log(x_1 + x_2 + \delta) - x_2^{1+\delta} \quad (17)$$

where $0 < \alpha_1 < 1 < \alpha_2$ with $A = \alpha_1 + \alpha_2 < 2$ and assuming that $\delta > 0$ is a sufficiently small parameter, implying that the game Γ_δ is indeed a strict social purpose game.

We investigate Nash equilibria as well as social optima of this class of strict social purpose games Γ_δ for $\delta > 0$ that are sufficiently small.

Nash equilibria: The game Γ_δ has a unique Nash equilibrium that can be determined by investigating the first order conditions for an interior solution. We derive that for $i = 1, 2$:

$$\frac{\partial \pi_i}{\partial x_i} = 0 \quad \text{implies} \quad \frac{\alpha_i}{x_1 + x_2 + \delta} = (1 + \delta)x_i^\delta$$

For the case that δ is sufficiently small, $\alpha_1 < 1$ and $\alpha_2 > 1$, it follows that there is no interior Nash equilibrium and that the only Nash equilibrium is the corner equilibrium given by $x_1^{\text{NE}} = 0$ and $x_2^{\text{NE}} = 1$.

Social optimum: The social optimum can be determined by optimising the standard social welfare function given by

$$\pi_1(x_1, x_2) + \pi_2(x_1, x_2) = A \log(x_1 + x_2 + \delta) - x_1^{1+\delta} - x_2^{1+\delta}$$

The first order conditions for this optimum can be derived and lead to the conclusion that

$$0 < x_1^{\text{SO}} = x_2^{\text{SO}} < \left(\frac{A}{2(1+\delta)} \right)^{\frac{1}{1+\delta}} < 1$$

due to the property that $A < 2$ and that $\delta > 0$ is sufficiently small.

Hence, in this game we have derived that $x_1^{\text{NE}} = 0 < x_1^{\text{SO}}$ as well as $x_2^{\text{NE}} = 1 > x_2^{\text{SO}}$, while $x_1^{\text{NE}} + x_2^{\text{NE}} = 1 < x_1^{\text{SO}} + x_2^{\text{SO}}$. \blacklozenge

It may happen that $x_i^{\text{NE}} < x_i^{\text{SO}}$ for any $i \in N$, as is shown in the next example.

Example 3.20 Consider the simple strict social purpose game Γ^s with players $N = \{1, \dots, n\}$ and $\bar{Q} > 0$ sufficiently large.⁵ For $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$ the payoff for player $i \in N$ is given by

$$\pi_i(x) = \alpha_i \sum_{i \in N} x_i - x_i^2 \quad \text{with } \alpha_i > 0 \quad i \in N. \quad (18)$$

Clearly Γ^s is a strict social purpose game, which unique Nash equilibrium is given by

$$x_i^{\text{NE}} = \frac{1}{2}\alpha_i \quad \text{resulting in } X^{\text{NE}} = \sum_{i \in N} x_i^{\text{NE}} = \frac{A}{2} \quad \text{with } A = \sum_{i \in N} \alpha_i. \quad (19)$$

The social optimum is determined as

$$x_i^{\text{SO}} = \frac{1}{2}A \quad \text{resulting in } X^{\text{SO}} = \sum_{i \in N} x_i^{\text{SO}} = \frac{n}{2}A. \quad (20)$$

Clearly, $\sum_{i=1}^n x_i^{\text{NE}} \leq \sum_{i=1}^n x_i^{\text{SO}}$ and, moreover, $x_i^{\text{NE}} < x_i^{\text{SO}}$ for every $i \in N$.

Now, for every $i \in N$, the Nash equilibrium payoffs are determined as $\pi_i(x^{\text{NE}}) = \frac{\alpha_i}{4}(2A - \alpha_i) > 0$ and the socially optimal payoffs are given by $\pi_i(x^{\text{SO}}) = \frac{A}{4}(2\alpha_i n - A)$. Note that $\pi_i(x^{\text{SO}}) > 0$ if and only if $\alpha_i > \frac{A}{2n}$. Finally, $\pi_i(x^{\text{NE}}) < \pi_i(x^{\text{SO}})$ if and only if $\frac{A}{\alpha_i} - \frac{\alpha_i}{A} < 2(n-1)$. \blacklozenge

⁵Hence, the strategic variable x_i can be assumed as if taken from \mathbb{R}_+ for every $i \in N$. Given the chosen formulation of the payoff structure, it is sufficient to select $\bar{Q} = A = \sum_{i \in N} \alpha_i > 0$.

4 Endogenous emergence of collaboration

Following the literature on partial cooperation in non-cooperative games developed in Chakrabarti et al. (2011, 2018, 2021) and Mallozzi and Tijs (2008, 2009, 2012), we consider the emergence of stable partial cooperative in social purpose games. We explore particularly the notion of a partial cooperative leadership equilibrium (PCLE) for this class of games prior to developing a stability notion and showing that there indeed might emerge stable partnerships in social purpose games under the PCLE concept. Through the PCLE notion and the definition of stable coalition formation, this incorporates a limited form of farsightedness in the formation of a stable coalition of cooperators.

4.1 Partial cooperative leadership equilibrium

We first consider two main notions of partial cooperative equilibrium, following the theory developed in Mallozzi and Tijs (2009) and Chakrabarti et al. (2011) as well as some of the insights obtained in Chakrabarti et al. (2018).

Within the context of a social purpose game Γ we have assumed throughout that players are ranked by their appreciation of the common benefit, i.e., $0 < \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$. This implies that the higher ranked players have a higher propensity to prefer the common benefit component of their payoffs. Hence, the higher ranked players have an increased preference for the benefit generated through cooperation or collaboration. This leads us to conclude that any cooperation would only emerge among the higher ranked players.

Using this as a fundamental assumption in any cooperation, we denote by the variable $k \in \{2, \dots, n\}$ a *level of cooperation*, which refers to the strategic willingness of the k highest ranked players to cooperate and to select a collective strategy for common purpose.⁶ Consequently, for any level of cooperation k , we denote by $C_k = \{n - k + 1, \dots, n\}$ the corresponding coalition of cooperators and by $N_k = N \setminus C_k = \{1, \dots, n - k\}$ the complement of C_k , being the corresponding set of non-cooperators.

We denote by $x^{C_k} = (x_{n-k+1}, \dots, x_n) \in \prod_{i=n-k+1}^n S_i = [0, \bar{Q}]^{C_k}$ and $x^{N_k} = (x_1, \dots, x_{n-k}) \in \prod_{j=1}^{n-k} S_j = [0, \bar{Q}]^{N_k}$. In particular, x^{C_k} is collectively determined, while x^{N_k} is competitively selected by individual players $j \in N_k$.

Equilibrium concepts under partial cooperation In this setting there naturally result two types of equilibrium concepts to be considered for any given level of cooperation $k \in \{2, \dots, n\}$. The first one is the *partial cooperative equilibrium* concept—based on the same logic as temporal “best response” rationality on which the standard Nash equilibrium concept is founded—in which every non-cooperator $i \in N_k = \{1, \dots, n - k\}$ acts competitively or individually, while all cooperating players $j \in C_k = \{n - k + 1, \dots, n\}$ act cooperatively as a single decision maker. Hence, in this partial cooperative equilibrium concept all decision makers in the set $\{1, \dots, n - k, C_k\}$ act as Nash optimisers. We refer to Chakrabarti et al. (2011, 2018) for a full analysis of partial cooperative

⁶We remark here that $k = 0$ or $k = 1$ are meaningless in the context of partial cooperation. Furthermore, the level of cooperation $k = n$ refers to a case of social optimality, since all players in the game act as cooperators. The latter is actually a potential valid level of cooperation.

equilibria in general non-cooperative games. Here we limit our discussion to the second notion of equilibrium under partial cooperation.

Alternatively, for any given level of cooperation $k \in \{2, \dots, n\}$, we might consider the notion of *partial cooperative leadership equilibrium* (PCLE). In such an equilibrium, it is assumed that the coalition of cooperators C_k acts as a single Stackelberg leader (von Stackelberg, 1934) and all non-cooperators $j \in N_k$ act as Stackelberg followers. In order to explicitly introduce this leader-follower equilibrium concept, for any $x^{C_k} \in [0, \bar{Q}]^{C_k}$ we denote by $\Gamma_k(x^{C_k}) = \langle N_k, [0, \bar{Q}]^{N_k}, \omega^{x^{C_k}} \rangle$ the normal form game—called the *conditional partial cooperative game* for x^{C_k} —given by player set $N_k = N \setminus C_k$ of non-cooperators whose strategy set is $S_j = [0, \bar{Q}]$ and who have a conditional payoff function $\omega_j^{x^{C_k}} : [0, \bar{Q}]^{N_k} \rightarrow \mathbb{R}$ defined by

$$\omega_j^{x^{C_k}}(x^{N_k}) = \pi_j(x^{N_k}, x^{C_k}) \quad j \in N_k = N \setminus C_k. \quad (21)$$

We denote by $\text{NE}(x^{C_k}) \subset [0, \bar{Q}]^{N_k}$ the set of Nash equilibria that emerge in this conditional non-cooperative game $\Gamma_k(x^{C_k})$.

Note that, under the assumptions of Proposition 3.5, $\text{NE}(x^{C_k}) \neq \emptyset$ but we cannot guarantee the uniqueness of the Nash equilibrium in $\Gamma_k(x^{C_k})$. Thus, in line with Mallozzi and Tijss (2009) and Chakrabarti et al. (2011), in order to select one Nash equilibrium among followers, we assume that cooperators are pessimistic in the sense that the coalition of the cooperators, the leader, supposes that the followers' (non-cooperators) choice is the worst for herself and select a maximin strategy. As a consequence, we introduce

$$\tilde{\pi}(x^{C_k}) = \min_{x^{N_k} \in \text{NE}(x^{C_k})} \sum_{i \in C_k} \pi_i(x^{N_k}, x^{C_k}) \quad (22)$$

and

$$\tilde{X}_k = \left\{ \tilde{x}^{C_k} \in [0, \bar{Q}]^{C_k} \mid \tilde{\pi}(\tilde{x}^{C_k}) = \max_{x^C \in [0, \bar{Q}]^C} \tilde{\pi}(x^C) \right\} \quad (23)$$

An action tuple $(x_*^{N_k}, x_*^{C_k}) \in [0, \bar{Q}]^{N_k} \times [0, \bar{Q}]^{C_k}$ is a *partial cooperative leadership equilibrium* for the game Γ if $x_*^{C_k} \in \tilde{X}_k$ and

$$x_*^{N_k} \in \arg \min_{x^{N_k} \in \text{NE}(x_*^{C_k})} \sum_{i \in C_k} \pi_i(x^{N_k}, x_*^{C_k}).$$

The following result states the conditions under which a PCLE exists in a social purpose game.

Proposition 4.1 *For every regular social purpose game Γ and every level of cooperation $k \in \{2, \dots, n\}$ there exists at least one partial cooperative leadership equilibrium.*

Proof. Assume that Γ is a regular social purpose game, i.e., H is concave and increasing, g_i are convex and continuous for all $i \in N$ and $h_i(x_i) = x_i$ for all $i \in N$. Take any level of cooperation $k \in \{2, \dots, n\}$ with the corresponding coalition of cooperators $C_k = \{n - k + 1, \dots, n\}$. We first show

the following claim:

Claim A: The correspondence $\mathbf{E}: [0, \bar{Q}]^{C_k} \rightarrow 2^{[0, \bar{Q}]^{N_k}}$ defined by

$$x^{C_k} \in [0, \bar{Q}]^{C_k} \mapsto \mathbf{E}(x^{C_k}) = \text{NE}(x^{C_k}) \subset [0, \bar{Q}]^{N_k} \quad (24)$$

is non-empty, compact valued and upper hemi-continuous.

Proof of Claim A: Since Proposition 3.5 holds, for any $x^{C_k} \in [0, \bar{Q}]^{C_k}$, $\text{NE}(x^{C_k}) \neq \emptyset$, implying that \mathbf{E} is non-empty valued.

Let $(x_p^{C_k})_{p \in \mathbb{N}}$ a sequence such that $x_p^{C_k} \rightarrow x^{C_k} \in [0, \bar{Q}]^{C_k}$. Let us take $x_p^{N \setminus C} \in \text{NE}(x_p^{C_k})$ such that $x_p^{N \setminus C} \rightarrow x^{N \setminus C}$ as $p \rightarrow \infty$. Since $x_p^{N \setminus C} \in \text{NE}(x_p^{C_k})$ we have

$$\begin{aligned} \alpha_j H \left(\sum_{j=1}^{n-k} x_{j,p}^{N_k} + \sum_{i=n-k+1}^n x_{i,p}^{C_k} \right) - g_j \left(x_{j,p}^{N_k} \right) \geq \\ \alpha_j H \left(x_j'^{N_k} + \sum_{l=1, l \neq j}^{n-k} x_{l,p}^{N_k} + \sum_{i=n-k+1}^n x_{i,p}^{C_k} \right) - g_j \left(x_j'^{N_k} \right) \end{aligned}$$

for every non-cooperator $j \in N_k$ and for any $x_j'^{N_k} \in [0, \bar{Q}]$.

If $p \rightarrow \infty$ by continuity of all functions we obtain

$$\begin{aligned} \alpha_j H \left(\sum_{j=1}^{n-k} x_j^{N_k} + \sum_{i=n-k+1}^n x_i^{C_k} \right) - g_j \left(x_j^{N_k} \right) \geq \\ \alpha_j H \left(x_j'^{N_k} + \sum_{l=1, l \neq j}^{n-k} x_l^{N_k} + \sum_{i=n-k+1}^n x_i^{C_k} \right) - g_j \left(x_j'^{N_k} \right) \end{aligned}$$

for every non-cooperator $j \in N_k$ and for any $x_j'^{N_k} \in [0, \bar{Q}]$. Thus, $x^{N_k} \in \text{NE}(x^{C_k})$ leading to the conclusion that \mathbf{E} is closed valued.

Finally, the set $\text{NE}(x^{C_k})$ is closed and compact for any $x^{C_k} \in [0, \bar{Q}]^{C_k}$, since $\text{NE}(x^{C_k}) \subset [0, \bar{Q}]^{N_k}$ is closed as shown above and $[0, \bar{Q}]^{N_k}$ is compact. Finally, since every closed correspondence with compact co-domain is upper hemi-continuous, this implies the assertion of the claim. \square

From Claim A and the fact that each payoff function π_i for $i \in C_k$ and π_j for $j \in N_k$ is continuous and quasi-concave, with reference to the proof of Chakrabarti et al. (2011, Theorem 2.6), there exists at least one PCLE in Γ . This shows the assertion of Proposition 4.1. \blacksquare

Remark 4.2 Denoting by $x_*^L = (x_*^{N_k}, x_*^{C_k})$ a partial cooperative leadership equilibrium at level of cooperation $k \in \{2, \dots, n\}$ and given $C_k \subset N$ as the corresponding coalition of cooperators, in the same assumptions of Proposition 3.16, we obtain that $\sum_{i \in C_k} x_i^{\text{NE}} \leq \sum_{i \in C_k} x_{*,i}^L$. \blacklozenge

The next example shows explicitly that the mapping \mathbf{E} as constructed in the proof of Proposition 4.1 does not have to be single-valued.

Example 4.3 Consider the regular social purpose game Γ with $n = 2$, $\bar{Q} = 1$ and payoff functions π_1 and π_2 given by choosing $\alpha_1 = \alpha_2 = 1$, $H(t) = t$, $g_1 = x_1$, $g_2 = 0$. Then $\pi_1(x_1, x_2) = x_2$ and $\pi_2(x_1, x_2) = x_1 + x_2$.

By considering the player 2 as the only cooperator, for any given x_2 , the other player maximised $\pi_1(x_1, x_2) = x_2$ with respect to x_1 . The multi-valued map E of Claim A assigns to every x_2 the whole strategy set $[0, 1]$. The non-cooperative player 1 determines her optimal strategy $\bar{x}_2 = 1$ by solving the problem

$$\max_{x_2} \left[\min_{x_1 \in [0,1]} x_1 + x_2 \right]$$

Hence, the map E is indeed multi-valued. ♦

4.2 Formation of stable coalitions of cooperators

Given the existence of a partial cooperative leadership equilibrium, the next objective is to address the following research question: is there a way to endogenously determine the number of cooperators in a partial cooperative framework?

For this, we introduce additional hypotheses to guarantee that a standard vNM stability conception (von Neumann and Morgenstern, 1947) can be implemented. In particular, it is required that in the game under consideration, there is a unique PCLE for every level of cooperation. We already established in Proposition 4.1 that every regular social purpose game admits at least one PCLE for every level of cooperation. Furthermore, for every strict social purpose game the constructed mapping E in the proof of Proposition 4.1 is a single-valued function. Therefore, we limit ourselves to the following class of social purpose games throughout the following discussion.

Axiom 4.4 *Throughout the following discussion we only consider strict social purpose games Γ that admit a unique Nash equilibrium x^{NE} as well as a unique partial cooperative leadership equilibrium (PCLE) for every level of cooperation $k \in \{2, \dots, n\}$.*

For every level of cooperation $k \in \{2, \dots, n\}$, representing the case that $C_k = \{n - k + 1, \dots, n\}$ forms as the coalition of cooperators, we denote by $x_i^L(k) \in [0, \bar{Q}]$ the unique PCLE strategy for every $i \in N$ and by $\pi_i^L(k) = \pi_i(x^L(k))$ the resulting PCLE payoff of player $i \in N$.

Under Axiom 4.4, we use the following notion of coalition stability originally proposed for similar decision situations in D'Aspremont et al. (1983), which in turn was informed by the vNM stability notion seminally stated in von Neumann and Morgenstern (1947). This particular definition of coalitional stability simply means that no player inside the coalition has an incentive to abandon her membership of the coalition of cooperators (“internal” stability) and no player outside the coalition of cooperators has an incentive to join the coalition of cooperators (“external” stability).

In the applied formulation below for a level of cooperation k , this is only expressed for the two marginal players $n - k \in N_k$ and $n - k + 1 \in C_k$ where as before $C_k = \{n - k + 1, \dots, n\}$ is the corresponding coalition of cooperators under cooperation level k . This is founded on the specific role of these marginal players, due to their position in the ranking based on their preference for the common benefit expressed through their respective α -weights. This definition has already been

applied in the motivating case of the tragedy of the commons as set out in Section 2. Therefore, the formalisation of this notion of stability is as follows.

Definition 4.5 Consider a strict social purpose game Γ that satisfies Axiom 4.4. We refer to the level of cooperation $k \in \{2, \dots, n\}$ as

- (i) **Internally stable** in Γ if $\pi_{n-k+1}^L(k) \geq \pi_{n-k+1}^L(k-1)$, and
- (ii) **Externally stable** in Γ if $k = n$ or $\pi_{n-k}^L(k) \geq \pi_{n-k}^L(k+1)$

where for ease of notation we denote by $x^L(1) = x^{\text{NE}}$.

The level of cooperation $k^* \in \{2, \dots, n\}$ is **stable** if k^* is an internally as well as externally stable level of cooperation in Γ .

The next result states the necessary and sufficient conditions for stability in the formation of a stable coalition of cooperators in an arbitrary strict social purpose game satisfying Axiom 4.4.

Proposition 4.6 Let Γ be a strict social purpose game satisfying Axiom 4.4 and consider the level of cooperation $k \in \{2, \dots, n\}$. Denote by $q = n - k + 1 \in C_k$ the marginal cooperator and $r = n - k \in N_k$ the marginal non-cooperator.⁷ Then the following properties apply:

- (i) The level of cooperation $k = 2$ is stable if and only if

$$\frac{1}{\alpha_{n-1}} \left(g_{n-1} \left(x_{n-1}^L(2) \right) - g_{n-1} \left(x_{n-1}^{\text{NE}} \right) \right) \leq H \left(\sum_{i \in N} x_i^L(2) \right) - H \left(\sum_{i \in N} x_i^{\text{NE}} \right)$$

as well as

$$\frac{1}{\alpha_{n-2}} \left(g_{n-2} \left(x_{n-2}^L(2) \right) - g_{n-2} \left(x_{n-2}^L(3) \right) \right) \geq H \left(\sum_{i \in N} x_i^L(2) \right) - H \left(\sum_{i \in N} x_i^L(3) \right).$$

- (ii) For $k \in \{3, \dots, n-1\}$ it holds that the level of cooperation k is stable if and only if

$$\frac{1}{\alpha_q} \left(g_q \left(x_q^L(k) \right) - g_q \left(x_q^L(k-1) \right) \right) \leq H \left(\sum_{i \in N} x_i^L(k) \right) - H \left(\sum_{i \in N} x_i^L(k-1) \right)$$

as well as

$$\frac{1}{\alpha_r} \left(g_r \left(x_r^L(k) \right) - g_r \left(x_r^L(k+1) \right) \right) \geq H \left(\sum_{i \in N} x_i^L(k) \right) - H \left(\sum_{i \in N} x_i^L(k+1) \right).$$

- (iii) The level of cooperation $k = n$ is stable if and only if

$$\frac{1}{\alpha_1} \left(g_1 \left(x_1^{\text{SO}} \right) - g_1 \left(x_1^L(n-1) \right) \right) \leq H \left(\sum_{i \in N} x_i^{\text{SO}} \right) - H \left(\sum_{i \in N} x_i^L(n-1) \right).$$

⁷We remark that as a consequence $C_{k-1} = C_k \setminus \{q\}$ and $C_{k+1} = C_k \cup \{r\}$.

Proof. We only state the proof for $k \in \{3, \dots, n-1\}$. The proofs for the two extreme cases $k \in \{2, n\}$ are omitted, but follow immediately from the reasoning developed below.

We compare the payoffs of player $q = n - k + 1 \in C_k$ in the stability conditions if she cooperates with C_k or acts as a non-cooperator $q \in N_{k-1}$. Indeed, if she cooperates, she receives payoff

$$\begin{aligned} u_q(C) &= \pi_q^L(k) = \alpha_q H \left(\sum_{i \in N} x_i^L(k) \right) - g_q \left(x_q^L(k) \right) \\ &= \alpha_q H \left(\sum_{i \in C_{k-1}} x_i^L(k) + x_q^L(k) + \sum_{j \in N_k} x_j^L(k) \right) - g_q \left(x_q^L(k) \right) \end{aligned}$$

and, if she does not cooperate with C_k , she would receive

$$\begin{aligned} u_q(NC) &= \pi_i^L(k-1) = \alpha_q H \left(\sum_{i \in N} x_i^L(k-1) \right) - g_q \left(x_q^L(k-1) \right) \\ &= \alpha_q H \left(\sum_{i \in C_{k-1}} x_i^L(k-1) + x_q^L(k-1) + \sum_{j \in N_k} x_j^L(k-1) \right) - g_q \left(x_q^L(k-1) \right). \end{aligned}$$

Internal stability requires now that $u_q(C) \geq u_q(NC)$. This is equivalent to the first condition in the assertion.

For the external stability condition of the level of cooperation $k \in \{3, \dots, n-1\}$, we compare the payoffs of the marginal non-cooperator $r = n - k \in N_k$ if she cooperates with C_k to form $C_{k+1} = C_k \cup \{r\}$ or acts as a non-cooperator $r \in N_k$. Indeed, if she cooperates with C_k she receives payoff

$$\begin{aligned} u_r(C) &= \pi_r^L(k+1) = \alpha_r H \left(\sum_{i \in N} x_i^L(k+1) \right) - g_r \left(x_r^L(k+1) \right) \\ &= \alpha_r H \left(\sum_{i \in C_k} x_i^L(k+1) + x_r^L(k+1) + \sum_{j \in N_{k+1}} x_j^L(k+1) \right) - g_r \left(x_r^L(k+1) \right) \end{aligned}$$

and, if she does not cooperate with C_k and remains in N_k , she receives

$$\begin{aligned} u_r(NC) &= \pi_r^L(k) = \alpha_r H \left(\sum_{i \in N} x_i^L(k) \right) - g_r \left(x_r^L(k) \right) \\ &= \alpha_r H \left(\sum_{i \in C_k} x_i^L(k) + x_r^L(k) + \sum_{j \in N_{k+1}} x_j^L(k) \right) - g_r \left(x_r^L(k) \right) \end{aligned}$$

External stability now requires that $u_r(NC) \geq u_r(C)$, which is equivalent to the second condition of the assertion. ■

Stability in the social purpose game of Example 3.20 We determine the partial cooperative leadership equilibrium (PCLE) for any level of cooperation of the simple social purpose game Γ^s

considered in the Example 3.20. Again assume that all players are ranked in order of their preference for the collectively generated benefits with $0 < \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$. In line with the previously applied notation, denote for any non-empty coalition $\emptyset \neq M \subseteq N$ by $A_M = \sum_{h \in M} \alpha_h > 0$. Recall that the payoffs are given by

$$\pi_i(x) = \alpha_i \sum_{i \in N} x_i - x_i^2 \quad \text{for every player } i \in N.$$

and that $x_i^{\text{NE}} = \frac{1}{2}\alpha_i$ and $x_i^{\text{NE}} = \frac{1}{2}A_N$

Select any level of cooperation $k \in \{2, \dots, n\}$. As before, $C_k = \{n - k + 1, \dots, n\} \subset N$ be the coalition of k cooperators with the highest preference for the generated social benefits in this game. The non-cooperators are now the players in $N_k = N \setminus C_k = \{1, \dots, n - k\}$ with the lowest preference for the collectively generated benefit.

Next, we compute the resulting unique PCLE for the level of cooperation k . Every non-cooperator $j \in N_k$ selects a best response to all other players' strategies. Hence,

$$x_j^L(k) = \frac{1}{2}\alpha_j \quad \text{for every non-cooperator } j \in N_k.$$

Note that this best response is independent on the chosen strategy x_{C_k} of the coalition of cooperators C_k .

The cooperators in C_k determine their strategies collectively to maximise their collective payoff $\pi_{C_k} = \sum_{i \in C_k} \pi_i$, given $(x_j^L)_{j \in N_k}$. This is again independent of the chosen strategies of all players $j \in N_k$ and results into

$$x_i^L(k) = \frac{1}{2}A_{C_k} = \frac{1}{2} \sum_{i=n-k+1}^n \alpha_i \quad \text{for every cooperator } i \in C_k.$$

Hence, in the partial cooperative leadership equilibrium we have that

$$X^L(k) = \sum_{h \in N} x_h^L(k) = \frac{1}{2}A_{N_k} + \frac{k}{2}A_{C_k} \quad (25)$$

$$\pi_j^L(k) = \alpha_j \left(\frac{1}{2}A_{N_k} + \frac{k}{2}A_{C_k} \right) - \frac{1}{4}\alpha_j^2 \quad \text{for every } j \in N_k \quad (26)$$

$$\pi_i^L(k) = \alpha_i \left(\frac{1}{2}A_{N_k} + \frac{k}{2}A_{C_k} \right) - \frac{1}{4}A_{C_k}^2 \quad \text{for every } i \in C_k \quad (27)$$

From these payoff levels we deduce that for every non-cooperator $j \in N_k$: $\pi_j^{\text{NE}} < \pi_j^L(k)$ and that for every cooperator $i \in C_k$ it holds that $\pi_i^{\text{NE}} < \pi_i^L(k)$ if and only if $\frac{A_{C_k}}{\alpha_i} - \frac{\alpha_i}{A_{C_k}} < 2(k - 1)$.

We next turn to the question whether partial cooperation in this simple social purpose game Γ^s results in a stable coalition of cooperators.

Proposition 4.7 *Under the partial cooperative leadership equilibrium concept, a level of cooperation*

$k \in \{2, \dots, n-1\}$ in the game Γ^s is stable if and only if

$$\alpha_{n-k+1} \geq \frac{A_{C_k}}{1 + \sqrt{2(k-1)}} \quad (28)$$

as well as

$$\alpha_{n-k} \leq \frac{A_{C_k}}{\sqrt{2k}}. \quad (29)$$

Furthermore, $k = n$ is a stable level of cooperation in Γ^s to form $C_n = N$ to implement the social optimum if and only if

$$\alpha_1 \geq \frac{A_N}{1 + \sqrt{2(n-1)}} \quad (30)$$

Proof. Here, and in the following, for simplicity of notation in formulas, we let $q = n - k + 1$ and $r = n - k$ denote the marginal cooperator $q \in C_k$ and the marginal non-cooperator $r \in N_k$, respectively.

If player q cooperates with C_k , she receives a payoff of

$$u_q(C) = \pi_q^L(k) = \alpha_q \left(\frac{1}{2} A_{N_k} + \frac{k}{2} A_{C_k} \right) - \frac{1}{4} A_{C_k}^2$$

and if she does not cooperate with C_k she receives

$$u_q(NC) = \alpha_q \left(\frac{1}{2} A_{N_k} + \frac{1}{2} \alpha_q + \frac{k-1}{2} A_{C_{k-q}} \right) - \frac{1}{4} \alpha_q^2.$$

Internal stability requires now that $u_q(C) \geq u_q(NC)$. This is equivalent to

$$\frac{\alpha_q}{2} k A_{C_k} - \frac{1}{4} A_{C_k}^2 \geq \alpha_q \left(\frac{k-1}{2} (A_{C_k} - \alpha_q) + \frac{\alpha_q}{2} \right) - \frac{1}{4} \alpha_q^2$$

or

$$A_{C_k}^2 - \alpha_q^2 \leq 2\alpha_q A_{C_k} + 2k\alpha_q^2 - 4\alpha_q^2$$

or

$$A_{C_k}^2 - 2\alpha_q A_{C_k} + \alpha_q^2 = (A_{C_k} - \alpha_q)^2 \leq 2(k-1)\alpha_q^2$$

This is equivalent to

$$A_{C_k} - \alpha_q \leq \sqrt{2(k-1)} \alpha_q \quad (31)$$

which is equivalent to the condition (28) in the assertion.

If the marginal non-cooperator $r = n - k \in N_k$ cooperates with C_k to form C_{k+1} , she receives payoff

$$u_r(C) = \alpha_r \left(\frac{1}{2}(A_{N_k} - \alpha_r) + \frac{k+1}{2}(A_{C_k} + \alpha_r) \right) - \frac{1}{4}(A_{C_k} + \alpha_r)^2$$

and, if she does not cooperate and remains member of N_k , she receives

$$u_r(NC) = \pi_r^L(k) = \alpha_r \left(\frac{1}{2}A_{N_k} + \frac{k}{2}A_{C_k} \right) - \frac{1}{4}\alpha_r^2.$$

External stability now requires that $u_r(NC) \geq u_r(C)$, which is equivalent to

$$\frac{k}{2}\alpha_r A_{C_k} - \frac{1}{4}\alpha_r^2 \geq \alpha_r \left(\frac{k}{2}\alpha_r + \frac{k+1}{2}A_{C_k} \right) - \frac{1}{4}(A_{C_k} + \alpha_r)^2$$

or

$$A_{C_k}^2 + 2\alpha_r A_{C_k} \geq \alpha_r [2(k+1)A_{C_k} + 2k\alpha_r - 2kA_{C_k}] = \alpha_r (2A_{C_k} + 2k\alpha_r) \quad (32)$$

which is equivalent to $A_{C_k}^2 \geq 2k\alpha_r^2$ and, hence, (29).

Finally, $k = n$ is a stable level of cooperation if and only if $u_1(C) \geq u_1(NC)$. This is equivalent to $A_N - \alpha_1 \leq \sqrt{2(n-1)}\alpha_1$ as derived above. ■

References

- ACEMOGLU, D. AND M. K. JENSEN (2013): “Aggregative Comparative Statics,” *Games and Economic Behavior*, 81, 27–49.
- BERGSTROM, T. C., L. BLUME, AND H. A. VARIAN (1986): “On the Private Provision of Public Goods,” *Journal of Public Economics*, 29, 25–49.
- CHAKRABARTI, S., R. P. GILLES, AND E. A. LAZAROVA (2011): “Strategic Behavior under Partial Cooperation,” *Theory and Decision*, 71, 175–193.
- (2018): “Partial Cooperation in Strategic Decision Situations,” *Theory and Decision*, 85, 455–478.
- (2021): “Stability of Cartels in Multi-market Cournot Oligopolies,” *The Manchester School*, 89, 70–85.
- CHANDER, P. AND H. TULKENS (1997): “The core of an economy with multilateral environmental externalities,” *International Journal of Game Theory*, 26, 379–401.
- CORCHÓN, L. C. (1994): “Comparative statics for aggregative games the strong concavity case,” *Mathematical Social Sciences*, 28, 151–165.
- CORNES, R. AND R. HARTLEY (2005): “Asymmetric contests with general technologies,” *Economic Theory*, 26, 923–946.
- D’ASPROMONT, C., A. JACQUEMIN, J. J. GABSZEWICZ, AND J. A. WEYMARK (1983): “On the Stability of Collusive Price Leadership,” *Canadian Journal of Economics*, 16, 17–25.
- DE MIGUEL, V. AND H. XU (2009): “A Stochastic Multiple-Leader Stackelberg Model: Analysis, Computation, and Application,” *Operations Research*, 57, 1220–1235.

- DIAMANTOUDI, E. AND E. S. SARTZETAKIS (2006): “Stable IEA: An Analytical Approach,” *Journal of Public Economic Theory*, 8, 247–263.
- DUBEY, P., O. HAIMANKO, AND A. ZAPECHELNYUK (2006): “Strategic complements and substitutes, and potential games,” *Games and Economic Behavior*, 54, 77–94.
- FRISCHMANN, B. M., A. MARCIANO, AND G. B. RAMELLO (2019): “Tragedy of the commons after 50 years,” *Journal of Economic Perspectives*, 33, 211–228.
- GILLES, R. P., M. PESCE, AND D. DIAMANTARAS (2020): “The Provision of Collective Goods through a Social Division of Labour,” *Journal of Economic Behavior & Organization*, 178, 287–312.
- HARDIN, G. (1968): “The tragedy of the commons,” *Science*, 162, 1243–1248.
- JENSEN, M. K. (2010): “Aggregative Games and Best-Reply Potentials,” *Economic Theory*, 43, 45–66.
- KWON, O. S. (2006): “Partial International Coordination in the Great Fish War,” *Environmental and Resource Economics*, 33, 463–483.
- LLOYD, W. F. (1833): *Two Lectures on the Checks to Population: Delivered Before the University of Oxford, in Michaelmas Term 1832*, Oxford, UK: J.H. Parker Publishers.
- MALLOZZI, L. AND R. MESSALLI (2017): “Multi-Leader Multi-Follower Model with Aggregative Uncertainty,” *Games*, 8, 25.
- MALLOZZI, L. AND S. TIJS (2008): “Conflict and Cooperation in Symmetric Potential Games,” *International Game Theory Review*, 10, 245–256.
- (2009): “Coordinating Choice in Partial Cooperative Equilibrium,” *Economics Bulletin*, 29, 1–6.
- (2012): “Stackelberg vs Nash Assumption in Partially Cooperative Games,” *Czech Economic Review*, 6, 5–13.
- MCGUINTY, M. AND G. MILAM (2013): “Public good provisions by asymmetric agents: experimental evidence,” *Social Choice and Welfare*, 40, 1159–1170.
- MONDERER, D. AND L. S. SHAPLEY (1996): “Potential Games,” *Games and Economic Behavior*, 14, 124–143.
- OLMSTEAD, S. M. (2014): “Climate change adaptation and water resource management: A review of the literature,” *Energy Economics*, 46, 500–509.
- OSTROM, E. (1990): *Governing the Commons: The evolution of institutions for collective action*, Cambridge, UK: Cambridge University Press.
- SELTEN, R. (1970): “Aggregations theorie des linear-quadratischen Modells,” in *Preispolitik der Mehrproduktenunternehmung in der statischen Theorie*, ed. by R. Selten, Springer Verlag, 113–125.
- VIVES, X. (1990): “Nash equilibrium with strategic complementarities,” *Journal of Mathematical Economics*, 19, 305–321.
- VON NEUMANN, J. AND O. MORGENSTERN (1947): *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 3rd ed.
- VON STACKELBERG, H. (1934): *Marktform und Gleichgewicht*, Habilitation, University of Vienna, translated as *Market Structure and Equilibrium*, Damien Bazin, Lynn Urch, and Rowland Hill (translators), Springer Verlag, Berlin, 2011 ed.